24. D. Kardassis, M. Hadzopoulou-Cladaras, D. P. Ramji, R. Cortese, V. I. Zannis and C. Cladaras, *MCBiol.* **10**, 2653 (1990).
25. J. A. A. Ladias, M. Hadzopoulou-Cladaras, D. Kardassis, P. Cardot, J. Cheng, V. Zannis and C. Cladaras, *JBC* **267**, 15849 (1992).
26. S. Metzger, T. Leff and J. L. Breslow, *JBC* **265**, 9978 (1990).
27. E. H. Ludwig, B. Levy-Wilson, T. Knott, V. D. Blackhart and B. J. McCarthy, *Cell Biol.* **10**, 329 (1991).
28. A. R. Brooks, and B. Levy-Wilson, *MCBiol.* **12**, 1134 (1992).
29. B. Levy-Wilson, B. Paulweber, B. P. Nagy, E. H. Ludwig and A. R. Brooks, *JBC* **267**, 18735 (1992).
30. B. Paulweber, A. R. Brooks, B. P. Nagy and B. Levy-Wilson, *JBC* **266**, 21956 (1991).
31. B. Paulweber and B. Levy-Wilson, *JBC* **266**, 24161 (1991).
32. B. D. Blackhart, Z. Yao and B. J. McCarthy, *JBC* **265**, 8358 (1990).
33. D. R. Goring, J. Rossant, S. Clapoff, M. L. Breitman and L. C. Tsui, *Science* **235**, 456 (1987).
34. A. R. Brooks, B. P. Nagy, S. Taylor, W. S. Simonet, J. M. Taylor and B. Levy-Wilson, *MCBiol.* **14** (1994). In press.
35. R. A. McKnight, A. Shamay, L. Sankaran, R. J. Wall and L. Hennighausen, *PNAS* **89**, 6943 (1992).
36. S. L. Hofmann, D. W. Russell, M. S. Brown, J. L. Goldstein and R. E. Hammer, *Science* **239**, 1277 (1988).
37. X. C. Jiang, L. B. Agellon, A. Walsh, J. L. Breslow and A. Tall, *J. Clin. Invest.* **90**, 1290 (1992).
38. F. Grosveld, G. B. van Assendelft, D. R. Greaves and G. Kollias, *Cell* **51**, 975 (1987).
39. A. Schedl, L. Montoliu, G. Kelsey and G. Schütz, *Nature* **362**, 258 (1993).
40. P. Fraser, S. Pruzina, M. Antoniou and F. Grosveld, *Genes Dev.* **7**, 106 (1993).
41. W. S. Simonet, N. Bucay, S. J. Lauer and J. M. Taylor, *JBC* **268**, 8221 (1993).
42. C. A. Pinkert, D. M. Ornitz, R. L. Brinster and R. D. Palmiter, *Genes Dev.* **1**, 268 (1987).
43. B. J. Aronow, R. N. Silbiger, M. R. Dusing, J. L. Stock, K. L. Yager, S. S. Potter, J. J. Hutton and D. A. Wiginton, *MCBiol.* **12**, 4170 (1992).
44. J. Mirkovitch, M.-E. Mirault and U. K. Laemmli, *Cell* **39**, 223 (1984).
45. P. N. Cockerill and W. T. Garrard, *Cell* **44**, 273 (1986).
46. C. Bonifer, M. Vidal, F. Grosveld and A. E. Sippel, *EMBO J.* **9**, 2843 (1990).
47. V. C. Blasquez, M. Xu, S. C. Moses and W. T. Garrard, *JBC* **264** 21183 (1989).
48. L. Emorine, M. Kuehl, L. Weir, P. Leder and E. E. Max, *Nature* **304**, 447 (1983).
49. T. G. Parslow and D. K. Grannar, *Nature* **299**, 449 (1982).
50. T. G. Parslow and D. K. Granner, *NARes.* **11**, 4775 (1983).
51. M. F. Linton, R. V. Farese, Jr., G. Chiesa, D. S. Grass, P. Chin, R. E. Hammer, H. H. Hobbs and S. C. Young, *J. Clin. Invest.* **92**, 3029 (1993).
52. M. J. Callow, L. J. Stoltzfus, R. M. Lawn and E. M. Rubin, *PNAS* **91**, 2130 (1994).

Early Growth Response Protein 1 (Egr-1): Prototype of a Zinc-finger Family of Transcription Factors

ANDREA GASHLER[1] AND VIKAS P. SUKHATME[2]

Department of Medicine
Beth Israel Hospital and Harvard Medical School
Boston, Massachusetts 02215

I. Overview of Immediate-early Genes 191
II. Identification of *Egr-1* cDNA by Differential Screening 192
III. *Egr-1* Is Expressed in Response to Diverse Stimuli 193
A. Induction by Mitogens 193
B. Induction during Development and Differentiation 195
C. Induction by Tissue or Radiation Injury 197
D. Induction in Neuronal Signaling 197
IV. Proximal Events 199
A. Second Messengers 199
B. *Egr-1* Promoter Analysis 200
V. Distal Events 201
A. Characterization of the Egr-1 Protein Product 201
B. DNA-binding Activity of Egr-1 205
C. Structure–Function Analysis 210
D. Targets of Egr-1 Regulation 216
VI. *In Vivo* Role of Egr-1 219
VII. *Egr-1* Is Part of a Gene Family, Including the Wilms Tumor Suppressor Gene *WT1* 219
VIII. Conclusion and Future Perspectives 220
References 221

I. Overview of Immediate-early Genes

Extracellular signals in the form of soluble factors, matrix proteins, and adhesion molecules influence the proliferation and differentiation of eukaryotic cells. These long-term responses, mediated by changes in gene expression, are coupled to biochemical events occurring in the plasma mem-

[1] Present address: Ligand Pharmaceuticals, San Diego, CA 92121.
[2] To whom correspondence may be addressed.

brane and cytosol that follow ligand–receptor interactions or other changes in the extracellular milieu. The so-called immediate-early genes are the earliest downstream nuclear targets for these events. These genes are, by definition, induced in the absence of *de novo* protein synthesis. In particular, a subclass of these genes encodes transcription factors, and these products form the first step in a cascade of gene–protein interactions. Thus, immediate-early transcription factor genes serve as nuclear couplers of early cytoplasmic events to long-term alterations in gene expression.

At present, the best characterized members of this group include c-*fos*, c-*jun*, and *Egr-1*. In turn, each of these genes is a prototype for a family of closely related proteins. This review focuses on the *Egr* gene family and its most extensively characterized member, *Egr-1*, first identified as an immediate-early gene responsive to growth factors and various differentiation cues, later confirmed to be a transcriptional regulatory protein. Other reviews have focused on changes in gene expression during the cell cycle (*1*) and transcriptional responses to extracellular signals (*2–4*). As a group, immediate-early transcription factors have provided important insights into how cellular responses to diverse extracellular signals are mediated.

II. Identification of *Egr-1* cDNA by Differential Screening

One approach to identifying novel genes that play key roles in cellular growth control is to focus on transcripts whose expression is low in nondividing cells but is rapidly up-regulated in cells stimulated by mitogen. Using c-*fos* as a model of immediate-early gene induction, several groups used similar differential screening strategies to isolate novel genes induced without intervening protein synthesis. Specifically, the following criteria were applied in our screen for important regulators of the G_0–G_1 transition: (1) Transcripts should be induced rapidly by serum stimulation of quiescent fibroblasts; (2) the mitogenic induction should not be affected by inhibitors of protein synthesis, such as cycloheximide; (3) expression should be induced by a spectrum of mitogens in a wide variety of cell types; and (4) the genes should be highly conserved in evolution (*5, 6*). In particular, we pursued differential screening of a library from BALB/c 3T3 cells stimulated for 3 hours with serum in the presence of cycloheximide. Clones were identified that hybridized preferentially to cDNA from serum and cycloheximide-treated fibroblasts as compared to cDNA from quiescent cells. The immediate-early gene c-*fos* was reisolated by this protocol. In addition, mitogenic stimulation of a variety of cell types from different species induced a 3.4-kb transcript. This novel immediate-early gene, designated *Egr-1* (*5–7*), has been independently cloned by similar differential screening strategies by a number of groups: *NGFI-A* was isolated as a nerve growth factor-inducible transcript in rat pheochromocytoma PC12 cells (*8*); *zif268* was cloned from serum-stimulated BALB/c 3T3 fibroblasts (*9*); *tis8* was identified as a phorbol-inducible gene in 3T3 cells (*10*); the chicken homolog, *cef5*, was cloned as a v-*src*-inducible gene from chicken embryo fibroblasts (*11*); and gene 225 was identified as a T-cell-activated transcript (*12*). Through hybridization to a highly conserved domain of the *Drosophila* factor Krüppel, *Krox24* was isolated from serum-stimulated 3T3 cells (*13*).

III. *Egr-1* Is Expressed in Response to Diverse Stimuli

A. Induction by Mitogens

In response to mitogens such as growth factors, hormones, and the tumor promoter TPA (phorbol), *Egr-1* induction is universal. In addition, *Egr-1* is expressed in diverse physiological contexts in particular cell types. The broad spectrum of extracellular stimuli that induce *Egr-1* can be roughly subgrouped into four categories: (1) mitogens, (2) developmental or differentiation cues, (3) tissue or radiation injury, and (4) signals that cause neuronal excitation.

In every cell type examined, *Egr-1* expression is rapidly induced by mitogenic stimulation. For example, in quiescent 3T3 cells stimulated with fetal calf serum, *Egr-1* expression is seen as early as 10 minutes, peaks around 30 minutes, and decays rapidly thereafter, returning to basal levels by 3–4 hours. Purified growth factors such as platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), and epidermal growth factor (EGF) also stimulate *Egr-1* expression in fibroblasts (*5, 9, 13*). The kinetics of induction are similar to those of c-*fos*, but the magnitude of *Egr-1* induction is typically severalfold greater (*5*).

In addition to induction in fibroblasts, mitogenic stimulation of *Egr-1* has been described in a wide array of cell types, such as kidney and liver epithelial cells and lymphocytes. For example, *Egr-1* is induced in regenerating liver within 1 hour after partial hepatectomy (*14*); in serum-starved BSC-1 monkey kidney epithelial cells in response to the mitogen adenosine diphosphate, in serum-deprived rat hepatoma H35 cells stimulated with serum or insulin, and in human peripheral blood lymphocytes treated with phytohemagglutinin (*5*). *Egr-1* is also up-regulated by protein tyrosine kinases, whose activity is associated with transformation in culture and tumorigenesis in animals. *Egr-1* message levels increase when a temperature-sensitive variant of v-Src is shifted from the nonpermissive to the permissive

temperature. *Egr-1* is similarly induced by expression of a second tyrosine kinase, v-Fos (*15, 16*). Because protein-tyrosine kinase activity has been implicated in events promoting cell division, *Egr-1* may be an important component of the mitogenic signal.

An extremely tight correlation between *Egr-1* expression and B lymphocyte activation has been established (*17*). B lymphocytes express surface immunoglobulin that acts as receptor for antigen. While mature B cells are activated by cross-linking surface immunoglobulin with anti-μ antibodies and respond by proliferating, immature B-cells, such as the WEHI-231 cell line, respond to anti-μ by down-regulation of proliferation and eventually cell death. The *Egr-1* response in mature and immature B lymphocytes differs accordingly: *Egr-1* is rapidly and transiently induced in mature B cells cross-linked with anti-μ but not in WEHI-231 cells treated identically. However, *Egr-1* can be induced to respond in WEHI-231 cells exposed to lipopolysaccharide (LPS), a treatment that protects these cells from the antiproliferative effects of anti-μ (*17*). The participation of *Egr-1* in positive versus negative signaling through surface immunoglobulin may be mediated by differential methylation of the gene. *Egr-1* is hypermethylated in immature B cells and in the WEHI-231 line. When an *Egr-1* reporter is transfected (*18*) into the WEHI line, it can be activated by anti-μ in contrast to the endogenous gene. Most convincingly, endogenous *Egr-1* can be induced in WEHI-231 cells treated with the inhibitor of methylation, 5′-azacytidine (*18*).

Additional correlation of *Egr-1* induction with mitogenicity has been shown in studies (*19*) in rat kidney mesangial cells. Numerous vasoactive agents, including PDGF, vasopressin, serotonin, and angiotensin II, induce proliferations in these cells, correlating *Egr-1* mRNA and protein induction with cell proliferation.

Strong evidence for a role for *Egr-1* in proliferation also comes from studies with mouse skeletal muscle Sol8 cells (*20*). Although *Egr-1* message was induced in response to mitogenic stimuli (such as basic fibroblast growth factor, PDGF BB, and fetal calf serum), differentiative stimuli (insulin), and other agents that caused neither proliferation nor differentiation, Egr-1 protein could be detected only in response to mitogenic cues. Translation of *Egr-1* may be uncoupled from transcriptional induction, as was in fact suggested by earlier studies with human fibroblasts (*21*). Although interferons α and γ, tumor necrosis factors α and β, and epidermal growth factor induced *Egr-1* message levels to a similar extent, the amount of *Egr-1* translated varied with the mitogenicity of the inducing agent. Cao *et al.* (*21*) suggest that the mechanism of translational regulation may be through the phosphorylation of cap-binding protein (eIF-4E). Phosphorylation of this factor, which promotes cellular protein synthesis, is enhanced by the mitogenic

agents EGF and tumor necrosis factor (TNF) but not interferon (IFN) (*21*). Together these studies present an intriguing correlation between the translatability of *Egr-1* message and the strength of the mitogenic inducing signal. Given the translational block in Egr-1 production induced by insulin in Sol8 cells, any role for Egr-1 in differentiated muscle must assume a function for the abundantly expressed *Egr-1* message, perhaps within its 3′ UTR (*22*). In light of these results, the assumption that *Egr-1* mRNA levels correlate with protein levels implicit in many studies of *Egr-1* induction must be reexamined.

Finally, recent work (*23*) suggests a role for Egr-1 in the regulation of astrocyte growth. Endothelin 3 (ET-3), a potent growth regulator in these cells, stimulates Egr-1 and basic fibroblast growth factor expression. An antisense oligonucleotide to *Egr-1* blocked ET-stimulated thymidine uptake and bFGF gene transcription. Moreover, an antisense oligomer to the bFGF gene significantly blocked ET-stimulated thymidine incorporation. These studies point to a causal role for *Egr-1* induction in the proliferation of astrocytes and suggest that the bFGF gene may be a relevant physiological target gene.

B. Induction during Development and Differentiation

In the adult mouse, high levels of *Egr-1* mRNA are seen in brain, thymus, heart, muscle, and lung. In particular, the high level of expression in the brain is located in the cerebral cortex and hippocampus (*14*). Lower levels are detected in kidney, spleen, and most other tissues, with very low levels in liver (*6*; *13*, *14*). A similar pattern of expression has been observed in the adult rat: *Egr-1* is most abundant in brain and adrenal gland, and is also highly expressed in superior cervical ganglia and lung (*24*, *25*).

During development, a single *Egr-1* transcript is predominantly expressed in cortex, midbrain, and cerebellum; in bone, cartilage, and muscle; and at several sites of epithelial–mesenchymal interactions. Studies in the developing rat suggest a role for *Egr-1* in postnatal maturation of the brain: *Egr-1* levels are low in neonatal and early postnatal brain, but increase dramatically at later times and in the adult animal, with highest levels detected in the cortex (*25*). In the developing mouse, *Egr-1* expression in 14.5- and 17.5-day fetal skeleton parallels c-*fos* expression, suggesting a role for these coregulated genes in skeletal development. *Egr-1* expression is correlated with the onset of ossification (about day 14.5) and is localized to regions of the embryo undergoing substantial bone formation, including the membranous and alveolar bones of the head and the periosteal and endochondral ossification sites of the developing long bones (*26*). Like c-*fos*, *Egr-1* is expressed in cartilage at the articular surfaces of joints and in the interstitial

cells that lie in between these elements. In addition, high-level *Egr-1* expression is seen in developing striated muscle, showing a patchy distribution. Finally, it has been suggested that *Egr-1* may respond to signals that mediate epithelial–mesenchymal interactions during organogenesis: expression is localized to ectodermally derived cells of the inner root sheath in young whisker follicles, in the underlying mesenchymal component of developing salivary and nasal glands of the mouse, as well as the mesenchymal component of the developing tooth (*26*). This initial patterning during tooth organogenesis requires primary signals derived from the dental epithelium. Importantly, recent *in vitro* reconstitution experiments demonstrate that purified bone morphogenetic protein 4 (BMP-4) can substitute for dental epithelium in inducing morphogenetic changes in the mesenchyme and in up-regulating *Egr-1* expression (*27*). In summary, the developmental profile of *Egr-1* is consistent with a role for it in brain maturation, in skeletal development, and in response to epithelial–mesenchymal interactions.

In several cell types, a rise in *Egr-1* expression is correlated with differentiative processes, in particular in cardiac, neural, osteoblast, and monocytic differentiation. Differentiation of P19 embryonal carcinoma cells into cardiac muscle, or nerve and glial cells, is induced in the presence of dimethyl sulfoxide (DMSO) or retinoic acid, respectively. In response to either, a biphasic pattern of *Egr-1* expression is seen. A transitory increase after 3 days of treatment is followed by high sustained levels of *Egr-1* expression after 14 days in culture (*6*). The expression of *Egr-1* in adult heart and brain is consistent with its prolonged response, pointing to a role for it in these differentiated cell types (*6, 28*). Neuronal differentiation can also be modeled on the rat pheochromocytoma cell line PC12. Nerve growth factor (NGF) causes an initial mitogenic response in PC12 cells, followed by growth arrest and differentiation into sympathetic neuronlike cells with extended neurites. *Egr-1* responds rapidly to NGF in PC12 cells, as to other growth factors; however, the expression is not transient, remaining high for up to 6 days (*6, 8*).

Finally, retinoic acid induces the differentiation of rat calvarial preosteoblastic RCT-1 cells. *Egr-1* is induced rapidly and transiently by retinoic acid in RCT-1 cells or primary cultures of embryonal calvarial cells, but not in the most mature RCT-3 line, which already expresses many osteoblastic markers (*29*). These observations, together with the expression of *Egr-1* in developing bone and cartilage described above, support a role for Egr-1 in osteoblast differentiation (*26, 29*).

As described above, *Egr-1* induction has been correlated with the onset of differentiation in several cell types. In particular, monocytic differentiation of U-937 and HL-60 myeloid leukemia cells induces *Egr-1* expression (*30, 31*). Interestingly, dexamethasone, an inhibitor of monocytic differentiation, blocks the *Egr-1* induction (*31*). Recent exciting results with myeloid cells provide the first demonstration that *Egr-1* expression is *necessary* for differentiation (*32*). The human myeloid leukemia cell line HL-60 can be induced to differentiate along either macrophage or granulocyte lineages by treatment with phorbol or DMSO, respectively. *Egr-1* expression is seen exclusively on induction of macrophage differentiation in HL-60 cells and primary myeloblasts. *Egr-1* antisense oligomers added to the culture medium prevent macrophage differentiation, and constitutive expression of Egr-1 limits the differentiative capacity of HL-60 cells such that these multipotent cells can be longer be induced for granulocyte differentiation (*32*). These results convincingly demonstrate that Egr-1 expression is essential for and restricts differentiation along the macrophage lineage. Mapping of the human *EGR1*[3] gene to chromosomal locus 5q31.1 is particularly intriguing with respect to these studies. The human EGR1 locus has been localized to a 2.8-megabase region defined by overlapping chromosomal deletions from patients with therapy-related acute myeloid leukemia (*6, 33*). The suggestion that *Egr-1* is a myeloid tumor-suppressor gene is consistent with a role for *Egr-1* in promoting myelogenesis.

C. Induction by Tissue or Radiation Injury

In a third context, *Egr-1* is induced in response to tissue or radiation injury. Ischemic injury to the kidney results in alterations in epithelial cell polarity, tissue damage, and cell death. Restoration of differentiated function after ischemic injury sets the kidney apart from the heart and brain, two organs that are irreversibly damaged by oxygen deprivation. Ischemic injury to rat kidney followed by reoxygenation induces a transient 30-fold increase in *Egr-1* expression that does not require protein synthesis. Moreover, because induction requires reoxygenation, *Egr-1* is not induced by the injury *per se*, but may rather act in response to postischemic events to mediate the subsequent processes of cell differentiation or proliferation (*34*). A second example of *Egr-1* induction as a consequence of cell injury is the cellular response to X-ray irradiation. Ionizing radiation has pleiotropic effects, including growth arrest, the repair of damaged DNA, and proliferation. *Egr-1* responds by a transient induction within 0.5 to 3 hours of exposure to X-rays in the absence of protein synthesis (*35*).

D. Induction in Neuronal Signaling

Immediate-early genes, by analogy to their part in the mitogenic response, may also play an important role in stimulus-transcription coupling in neurons (*36*). Several lines of experimentation indicate that immediate-early

[3] EGR is the human factor; Egr is the mouse or rat factor.

Response	Stimulus	Cell type	References
Mitogenic	serum	fibroblasts	5, 6, 9, 13
	PDGF, EGF, FGF	fibroblasts	5, 6, 9, 13, 43
	insulin	hepatocytes	5
	phytohemagglutinin	peripheral blood lymphocytes	5, 12
	anti-μ	B lymphocytes	17
	adenosine diphosphate	kidney epithelial cells	5
	PDGF, vasopressin	kidney mesangial cells	19, 114
	serum	skeletal muscle Sol8 cells	20
	bFGF, PDGF BB	skeletal muscle Sol8 cells	20
	partial hepatectomy	liver	9, 119
	GM-CSF, LPS	peritoneal macrophages	121, 115
	endothelin	astrocytes	23
	angiotensin II	vascular smooth muscle cells	120
Hypertrophic	endothelin, angiotensin II	myocyte	101, 117
Differentiative	NGF	pheochromocytoma PC12 (neural)	6, 8
	retinoic acid, DMSO	embryonal carcinoma P19	6
	retinoic acid	embryonal calvarial cells, RCT-1 (osteoblast)	29
	TPA, DMSO	myeloid leukemia HL60 and U-937	30, 31, 32
Tissue/ radiation injury	ischemia	kidney	34
	ionizing radiation	293, SQ-20B	35, 53, 116
Neuronal excitation	potassium ions	PC12 (depolarization)	36, 6, 37
	metrazole	seizures in vivo	6
	NMDA	hippocampus	40
	visual stimuli	visual cortex	38, 39
	electroconvulsive shock therapy, dopamine receptor activation, opiate withdrawal	CNS	41
	peripheral nervous system	sciative nerve transection	42
Other	urea	MDCK, LLC-PK$_1$ renal epithelial cells	118

FIG. 1. Biological processes in which *Egr-1* expression has been described.

genes, including *Egr-1*, participate in the rapid response of neurons to *trans*-synaptic stimuli. *In vivo*, *Egr-1* levels increase rapidly in the brain following seizure activity, with kinetics similar to c-*fos* (6). Membrane depolarization of PC12 cells by treatment with potassium chloride also results in rapid and transient induction of *Egr-1* (6, 37). In dark-reared cats, a brief 1-hour visual stimulation causes dramatic and transient induction of *Egr-1*, c-*fos*, and *junB* mRNAs that are specific to the visual cortex, i.e., absent from the frontal cortex. The magnitude of the induction, greatest in young animals, is consistent with the idea that *Egr-1* expression plays a fundamental role during the critical period of development in the visual cortex (38, 39). A role for *Egr-1* in postnatal maturation of the brain is supported by the dramatic increase in *Egr-1* message levels in all sections of postnatally developing rat brain,

especially cortex (25). Finally, high-frequency stimulation of the perforant path–granule cell synapse results in induction of *Egr-1* in the postsynaptic cells. The response of *Egr-1* is highly reproducible, as compared to the variable response of other immediate-early genes. Interestingly, induction of *Egr-1* is correlated with long-term potentiation, because both responses require the *N*-methyl-D-aspartate receptor and a stimulus of similar frequency and intensity (40). Additional studies show *Egr-1* induction following electroconvulsive shock therapy, D1 dopamine receptor activation, and opiate withdrawal (41). Transient *Egr-1* induction has also been noted in the peripheral nervous system, e.g., sciatic nerve transection provokes Egr-1 protein increase in neurons of the spinal dorsal horn (42). These studies, and the expression of *Egr-1* in developing and adult brain and in the peripheral nervous system are consistent with a role for Egr-1 in neurophysiological processes.

This summary of the contexts in which *Egr-1* is expressed emphasizes the diversity of signals that induce *Egr-1* (Fig. 1). *Egr-1* is induced by mitogenic stimuli in all cell types; during differentiation of nerve, cardiac, bone, and myeloid cells; after tissue injury due to ischemia or irradiation; and by signals that result in neuronal excitation, such as membrane depolarization or brain seizures. There has been one demonstration, in the differentiation-inducible HL-60 cell line, of a phenotype resulting from inappropriate Egr-1 expression (32). In addition to promoting and restricting differentiation of myeloid precursors along the macrophage lineage, the enormous complexity of the Egr-1 response hints that this protein may play diverse roles in different cellular contexts.

IV. Proximal Events

A. Second Messengers

Two strategies have yielded insight into the complex regulation of the *Egr-1* gene: activation or inhibition of specific second-messenger pathways and a molecular genetic dissection of the *Egr-1* promoter. Multiple intracellular pathways appear to contribute to the regulation of *Egr-1* expression. Both protein-kinase-C (PKC)-dependent and -independent mechanisms are integral in linking extracellular signals to transcriptional activation of *Egr-1*. Clearly, the PKC pathway can relay extracellular stimuli to a nuclear response resulting in *Egr-1* induction, because direct activation of the pathway by phorbol ester (TPA) induces *Egr-1* (5, 43). In addition, non-PKC pathways also play a role: fibroblasts rendered deficient in PKC signaling by long-term exposure to phorbol retain a robust *Egr-1* response to serum and epidermal growth factor (43).

In the response of *Egr-1* to tumor necrosis factor and interferon in human fibroblasts, the PKC pathway appears instrumental. Treatment with H7 (a nonspecific inhibitor of protein kinases including PKC) or the PKC inhibitor staurosporine effectively blocks much of the *Egr-1* response. The selective inhibitor of cyclic-nucleotide-dependent protein kinases, HA1004, does not modify the *Egr-1* response (*21*). Stimulation of B lymphocytes with phorbol or the PKC agonist SC-9 also up-regulate *Egr-1* expression, implying that surface immunoglobulin (Ig)-generated signals work through PKC. Evidence for the PKC pathway as a requisite component of anti-μ induction of *Egr-1* comes from studies with inhibitors of PKC. A prior treatment with either H7 or sangivamycin, effective inhibitors of PKC, blocks the increase in *Egr-1* mRNA levels in response to anti-μ. Again, the cyclic-nucleotide-dependent protein kinase inhibitor HA1004 had no effect. These studies demonstrate that activation of PKC is involved in coupling surface Ig stimulation in B lymphocytes to the transcriptional response of the *Egr-1* gene (*44*).

The PKC pathway appears fundamental in mediating *Egr-1* induction in response to X-irradiation. First, prolonged stimulation with micromolar concentations of phorbol depletes PKC and virtually blocks the X-ray inducibility of Egr-1 in SQ20B cells. In addition, pretreatment with the inhibitor H7 but not HA1004 markedly attenuates the X-ray inducibility of *Egr-1* in SQ20B or 293 cells (*35*).

In contrast, an intracellular pathway involving c-Raf plays a central role in the v-Src induction of *Egr-1*. c-Raf-1 is a serine-threonine protein kinase, and v-Raf up-regulates the *Egr-1* promoter. Moreover, expression of a kinase-defective mutant of c-Raf-1 blocks induction of *Egr-1* upon regulation of the *Egr-1* gene.

B. Egr-1 Promoter Analysis

The architecture of the *Egr-1* promoter has been described by several groups who have cloned the murine (*14*, *46*), rat (*47*), and human *Egr-1* genes. In particular, the coregulation of c-*fos* and *Egr-1* in several contexts has prompted a comparison of their promoter sequences. Six CC(W)6GG elements (CA_rG boxes), the functional core of the serum response element (SRE), are present in the *Egr-1* promoter; however, none of these potential SREs shares the extended symmetry outside of the core sequence that typifies the c-*fos* SRE (*48*). In addition to the CA_rG boxes, putative regulatory elements in the *Egr-1* promoter include cAMP response elements, AP1, CREB, and Sp1 sites as well as a CCAAT box and TATA motif (*14*, *46*, *47*, *49*), as illustrated in Fig. 2.

The demonstration that 1 kb of murine 5′ sequence confers serum and phorbol responsiveness to a CAT reporter in mouse fibroblasts opened the door to delineation of the functional elements (*14*, *50*, *51*). Similarly, NGF inducibility was observed with the sequence from −532 to +100 of the rat gene in PC12 cells (*47*). Deletion analysis of the *Egr-1* promoter showed that a construct with sequence to −594 (and all six CA_rG boxes) retains full serum inducibility whereas deletion to −166 (with the two proximal CA_rG elements) has partial serum responsiveness as compared to a minimal promoter construct. Moreover, synthetic constructs with a single *Egr-1* CA_rG box confer serum inducibility on the heterologous thymidine kinase promoter (*49*). These results show clearly that the decanucleotide inner core of the previously defined c-*fos* SRE functions as a serum response element in the *Egr-1* promoter. In a gel-shift assay, the core *Egr-1* SRE can compete for binding against the c-*fos* SRE with its more extensive dyad symmetry. And like the c-*fos* SRE, the *Egr-1* CA_rG boxes bind to *in-vitro*-translated serum response factor.

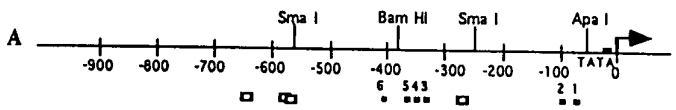
Further experiments with synthetic constructs indicate that tandem copies of the CA_rG boxes are more strikingly inducible than an individual element (*49*). Given these observations, the greater serum inducibility of *Egr-1* versus c-*fos* may be explained by the multiple elements in the *Egr-1* promoter as compared to the single SRE regulating c-*fos* expression. The CA_rG box appears to play a central role in the broad responsiveness of *Egr-1* to mitogens, because this motif directs induction by PDGF, phorbol, v-*src*, and v-*fos*, as well as serum (*16*, *49*, *50*, *52*). These elements, especially the three most 5′ ones, are also responsible for the activation of *Egr-1* by ionizing radiation (*53*).

Finally, the CA_rG boxes in the *Egr-1* mediate the down-regulation of *Egr-1* transcription following mitogenic stimulation. In particular, the Fos protein effects this transcriptional repression; Fos mutants lacking a leucine zipper function as well in this assay, and the C-terminal region of Fos is sufficient for this function (*51*).

V. Distal Events

A. Characterization of the Egr-1 Protein Product

Immediate-early genes encode several types of proteins, including growth factors, growth factor receptors, cytoskeletal proteins, and transcription factors. Sequence analysis of Egr-1 revealed a protein with three tandemly repeated Cys_2His_2 zinc-finger motifs that presage the function of this protein (*6*, *8*, *13*, *14*). The zinc finger (see below), a highly conserved eukaryotic DNA-binding motif, is a compact domain that uses conserved pairs of cysteine and histidine residues to coordinate a central zinc ion (*54*). The importance of the *Egr-1* gene product is also suggested by the conservation



▪ serum response element (SRE)
▫ Egr-1 binding site (EBS)

B

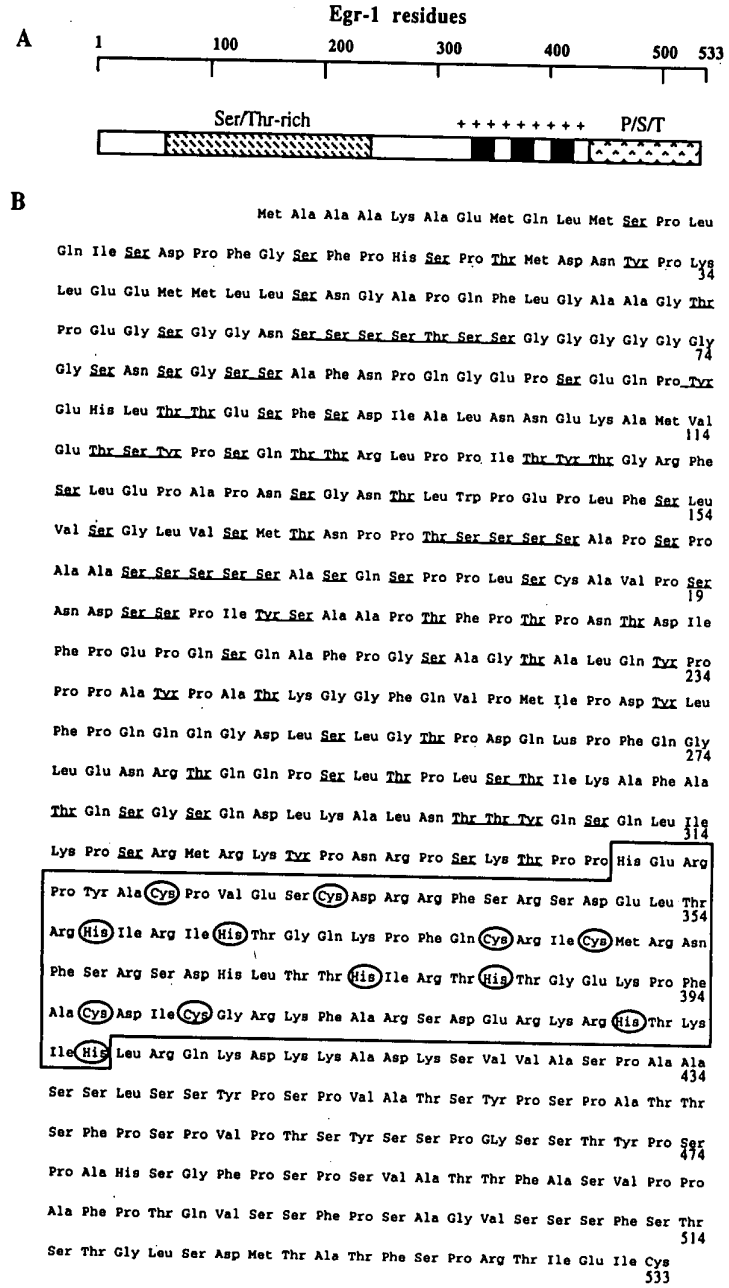
```
ACGGAGGGAA TAGCCTTTCG ATTCTGGGTG GTGCATTGGA AGCCCCAGGC TCTAAAACCC
-935                                                          -876

CCAACCTACT GACTGGTGGC CGAGTATGCA CCCGACTGCT AGCTAGGCAG TGTCCCAAGA
-875                                                          -816

ACCAGTAGCC AAATGTCTTG GCCTCAGTTT TCCCGGTGAC ACCTGGAAAG TGACCCTGCC
-815                                                          -756

ATTAGTAGAG GCTCAGGTCA GGGCCCCGCC TCTCCTGGGC GGCCTCTGCC CTAGCCCGCC
-755                                                          -696

CTGCCGCTCC TCCTCTCCGC AGGCTCGCTC CCACGGTCCC CGAGGTGCGC GGGTGAGCCC
-695                                                          -636

AGGATGACGG CTGTAGAACC CCGGCCTGAC TCGCCCTCGC CCCCGCGCCG GGCCTGGGCT
-635                                                          -576

TCCCTAGCCC AGCTCGCACC CGGGGGCCGT CGGAGCCGCC GCGCGCCCAG CTCTACGCGC
-575                                                          -516

CTGGCCCTCC CCACGCGGGC GTCCCCGACT CCCGCGCGCG CTCAGGCTCC CAGTTGGGAA
-515                                                          -456

CCAAGGAGGG GGAGGATGGG GGGGGGGGTG TGCGCCGACC CGGAAACGCC ATATAAGGAG
-455                                                          -396

CAGGAAGGAT CCCCCGCCGG AACAGACCTT ATTTGGGCAG CGCCTTATAT GGAGTGGCCC
-395                                                          -336

AATATGGCCC TGCCGCTTCC GGCTCTGGGA GGAGGGGCGA GCGGGGGTTG GGCGGGGGC
-335                                                          -276

AAGCTGGGAA CTCCAGGCGC CTGGCCCGGG AGGCCACTGC TGCTGTTCCA ATACTAGGCT
-275                                                          -216

TTCCAGGAGC CTGAGCGCTC GCGATGCCGG AGCGGGTCGC AGGGTGGAGG TGCCCACCAC
-215                                                          -156

TCTTGGATGG GAGGGCTTCA CGTCACTCCG GGTCCTCCCG GCCGGTCCTT CCATATTAGG
-155                                                          -96

GCTTCCTGCT TCCCATATAT GGCCATGTAC GTCACGGCGG AGGCGGGCCC GTGCTGTTCC
-95                                                           -36

AGACCCTTGA AATAGAGGCC GATTCGGGGA GTCGCGAGAG ATCCCAGCGC GCAGAACTTG
-35                                                           +25

GGGAGCCGCC GCCGCGATTC GCCGCCGCCG CCAGCTTCCG CCGCCGCAAG ATCGGCCCCT
+26                                                           +85

GCCCCAGCCT CCGCGGCAGC CCTGCGTCCA CCACGGGCCG CGGCTACCGC CAGCCTGGGG
+86                                                           +145
```


of the coding sequence across vertebrate evolution: human (7), rat (8), mouse (6, 13, 14), chicken (11), and zebrafish (55, 56) cDNAs are highly homologous.

From the deduced amino-acid sequence of Egr-1 protein, several interesting features have been predicted (Fig. 3A). Basic residues cluster in the three zinc fingers and adjacent sequence. The amino-terminal 300 amino acids are rich in proline (14%) and serine/threonine residues (24%). Several stretches of five to seven consecutive serine or threonine residues are present with one series of seven serine/threonine residues followed by seven glycines (Fig. 3B). It has been noted that the repeating trinucleotide motifs that encode these poly(aminoacid) stretches are similar to those whose expansion has been implicated in human disease (55, 57, 58). The region on the carboxy-terminal side of the zinc-finger motifs is also rich in proline (15%) and serine plus threonine (37%), but this region is distinguished by a repeated motif of eight amino acids with the consensus Ser/Thr-Ser/Thr-Phe/Tyr-Pro-Ser-Pro-X-X. The composition of this reiterated sequence is reminiscent of the heptapeptide repeat in the carboxy-terminal domain of the RNA polymerase II large subunit (59). The proline-rich regions of Egr-1 are predicted to lack α-helical secondary structure, whereas the high content of serine, threonine, and tyrosine residues suggests that Egr-1 may be phosphorylated.

Characterization of the *Egr-1* gene product showed it to encode a short-lived protein with an anomalous electrophoretic mobility of 80–82 kDa. In fibroblasts, Egr-1 protein is rapidly induced by serum, accumulating within 30 minutes and reaching maximum levels at 1–2 hours poststimulation (50). Consistent with its putative DNA-binding function, immunocytochemistry and cell fractionation studies show that Egr-1 is located in the nucleus (50, 60, 61). Studies (60) have characterized the rat homolog in PC12 cells with several antisera directed against various regions of the protein. In particular, a truncated species of 54 kDa is cytoplasmic. This 54-kDa species is recognized by antisera directed against the basic region immediately 5′ of the first zinc finger but not by sera against a C-terminal peptide. These results were an early indication that sequences within or C-terminal to the zinc-finger

FIG. 2. The 5′ upstream sequence of the murine *Egr-1* gene. (A) Schematic of promoter depicting putative regulatory elements. The positions of the six serum response elements within approximately 1 kb of promoter sequence are depicted as darkened boxes. The locations of the Egr-1 binding sites within the promoter are indicated as open boxes. (B) Nucleotide sequence of the *Egr-1* promoter. [Reprinted from *NARes* (Ref. 46) by permission of Oxford University Press.] The nucleotides are numbered from the cap site, which is +1. The putative TATA element is underlined and three Egr-1 binding sites in the 5′ promoter region are boxed.



domain may participate in nuclear targeting (*60*). The *Egr-1* gene product is also phosphorylated: alkaline phosphatase converts the two closely spaced Egr-1 species seen on SDS–PAGE analysis of NGF-stimulated PC12 cells to the faster migrating form (*50*, *60*). Immunoprecipitation of Egr-1 from phosphate-labeled HeLa cells and subsequent analysis of phosphoaminoacid content indicate that the phosphorylation is on serine (*62*).

B. DNA-binding Activity of Egr-1

Hundreds of eukaryotic transcription factors share the highly conserved DNA-binding motif known as the zinc finger. First identified as a compact zinc-binding domain in the *Xenopus* transcription factor IIIA (*54*), this well-conserved motif also occurs in the yeast proteins SW15 and ADR1, *Drosophila* factor Krüppel and Hunchback, and mammalian regulatory proteins such as the testis-determining factor ZFY, the enhancer-binding protein Sp1, and the Wilms tumor suppressor WT1. TFIIIA-like fingers are distinguished by pairs of conserved cysteine and histidine residues and are evolutionarily and structurally distinct from the cysteine-rich zinc-binding motifs in the steroid receptors and in the yeast factor GAL4. A variable number of tandem repeats of this domain of 28–30 amino acids act in concert to recognize a specific DNA sequence (*63*).

Residues fundamental to the structural integrity of the finger domain are conserved among all Cys_2His_2 zinc-finger proteins whereas other amino acids involved in base sequence discrimination may be unique or confined to a subset of this large family of proteins. Pairs of cysteine and histidine residues are absolutely conserved as are usually the hydrophobic amino acids phenylalanine and leucine (Fig. 4A). The region connecting the histidine of one finger to the cysteine of the following finger, designated the H–C link, has the highly conserved consensus His-Thr-Gly-Glu-Lus/Arg-Pro-Phe-Tyr-X-Cys (*63*). In addition, three variable residues, discussed below, appear to participate in sequence-specific interactions with DNA.

NMR and cystallographic studies suggest that each zinc-finger motif consists of an antiparallel β-sheet that includes the two consensus cysteines, and an α-helix that contains the two conserved histidine residues (Fig. 4B). Each

FIG. 3. Schematic structure and amino-acid sequence of the Egr-1 protein. (A) Structural features of Egr-1. Each zinc-finger motif is designated by a black bar. The basic region of Egr-1 is indicated (+++). The serine/threonine-rich N-terminal domain of Egr-1 is shown on the left and the proline/serine/threonine-rich C-terminus (P/S/T) is on the right. (B) Coding sequence of murine Egr-1. [Reprinted with permission from Ref. 6, copyright 1988 Cell Press.] The three zinc-finger motifs are enclosed. Conserved cysteine and histidine residues in the zinc fingers are circled. Serine, threonine, and tyrosine residues in the N-terminal domain are underlined.

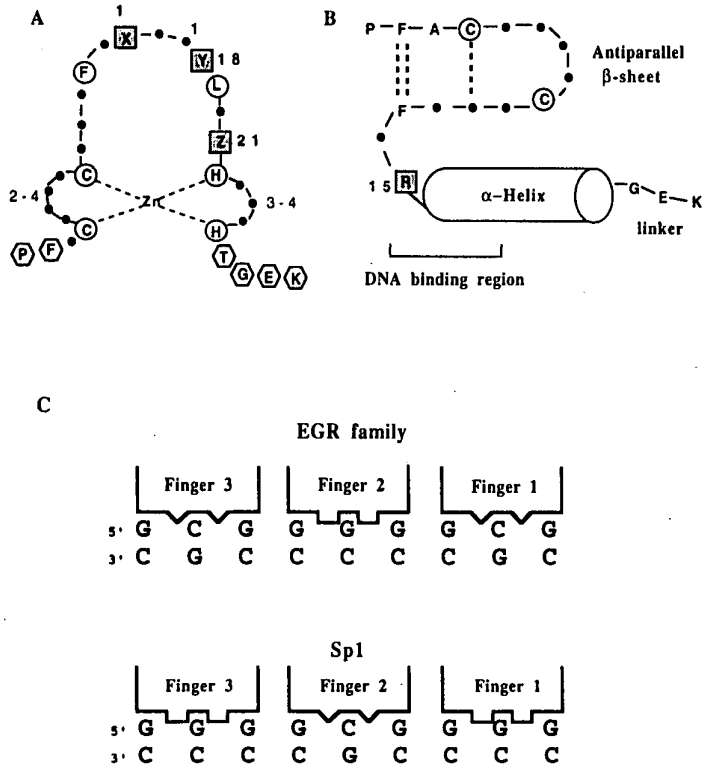


FIG. 4. The zinc finger as a modular DNA-binding motif. (A) Zinc-finger consensus residues. Invariant cysteine (C) and histidine (H) residues that coordinate a zinc ion are circled, as are the conserved hydrophobic residues phenylalanine and leucine. Residues that are part of the highly conserved His–Cys link are enclosed by hexagons. Amino acids that determine the sequence-specificity of binding are shown in the shaded boxes. (B) Diagram of zinc-finger folding (from *113*). Each zinc-finger domain is composed of a β-sheet and an α-helix. Hydrogen bonds are depicted as dotted lines. (C) Each zinc finger contacts a three-nucleotide subsite. [Reprinted with permission from *Nature* (Ref. 66), copyright 1991 Macmillan Magazines Limited.] Fingers 1 and 3 of Egr-2 are postulated to bind the same three-nucleotide subsite as finger 2 of Sp1.

zinc finger incorporates these secondary structures into a compact globular domain with the invariant cysteine and histidine residues coordinating a central zinc ion. A hydrophobic core including the conserved phenylalanine and leucine residues and the first histidine stabilizes the domain. In a manner similar to prokaryotic helix–turn–helix motifs and eukaryotic homeodomains, the α-helix of the zinc finger lies within the major groove of DNA. Multiple interactions between amino-acid side-chains of the helix and DNA base-pairs combine to discriminate among nucleic-acid sequences (*64*).

In a search for the DNA element recognized by Egr-1 among fragments derived from the 5′ upstream flanking sequence of the *Egr-1* gene (*49*), it was found that autoregulation by other immediate early genes such as c-*fos* and cycloheximide superinduction of *Egr-1* was consistent with the hypothesis that *Egr-1* regulates its own expression. Using gel mobility shift assays with Egr-1 protein purified from bacteria, specific binding to one promoter fragment was observed (*49*). DNase-I footprinting identified the sites of contact, revealing that Egr-1 binds the 9-bp sequence GCG-GGG-GCG. Further gel shifts comparing the affinity of this sequence to sites altered at various positions generated the consensus sequence: GCG-KGG-GCG (*49*). Gel shift assays with zinc-chelating agents were utilized to demonstrate the requirement for zinc cations to effect DNA binding (*50*).

Two types of experiments support a similar model for the determination of DNA-binding specificity by EGR fingers and proteins with related zinc-finger domains (reviewed in *65*). Mutagenesis experiments were guided by the similar but distinct zinc-finger domains of Sp1 and Egr-2, a gene whose three zinc fingers are identical to those of Egr-1 except for four conservative amino-acid substitutions. These mutagenesis studies foreshadowed the results obtained by cocrystallization of the Egr-1 zinc-finger domain and its cognate binding site. It has been observed that Sp1 and Egr-2 each contain three zinc fingers and bind to a (G+C)-rich 9-bp binding site (*66*). If each motif interacts with DNA in an analogous manner, then a zinc finger is predicted to contact 3 bp of DNA. Furthermore, comparison of the Egr consensus GCG-GGG-GCG to the Sp1 consensus GGG-GCG-GGG suggested that fingers 1 and 3 of Egr-2 might have the same specificity-determining residues as finger 2 of Sp1 (Fig. 4C). Fingers 1 and 3 of Egr-2 share Glu^{18} and Arg^{21} with finger 2 of Sp1 (Fig. 5A). In addition, finger 2 of Egr-2 and fingers 1 and 3 of Sp1 each have a histidine residue at position 18 of the finger. It was predicted (*66*) that the residues at positions 18 and 21 discriminate between GGG or GCG subsites. In accordance with this hypothesis, mutagenesis of Egr-2 finger 2 residues His^{18} to Glu and Thr^{21} to Arg created a protein that did not bind the *Egr-2* cognate sequence but instead recognized the novel sequence GCG-GCG-GCG (*66*).

The mode thus constructed (*66*), in which variable residues at positions 18 and 21 were postulated to be the determinants of base-sequence discrimination, has been substantiated (*64*). Solution of the Egr-1 zinc-finger domain–DNA crystal structure provided a framework for understanding how proteins with tandemly repeated Cys_2His_2 zinc fingers interact with DNA.

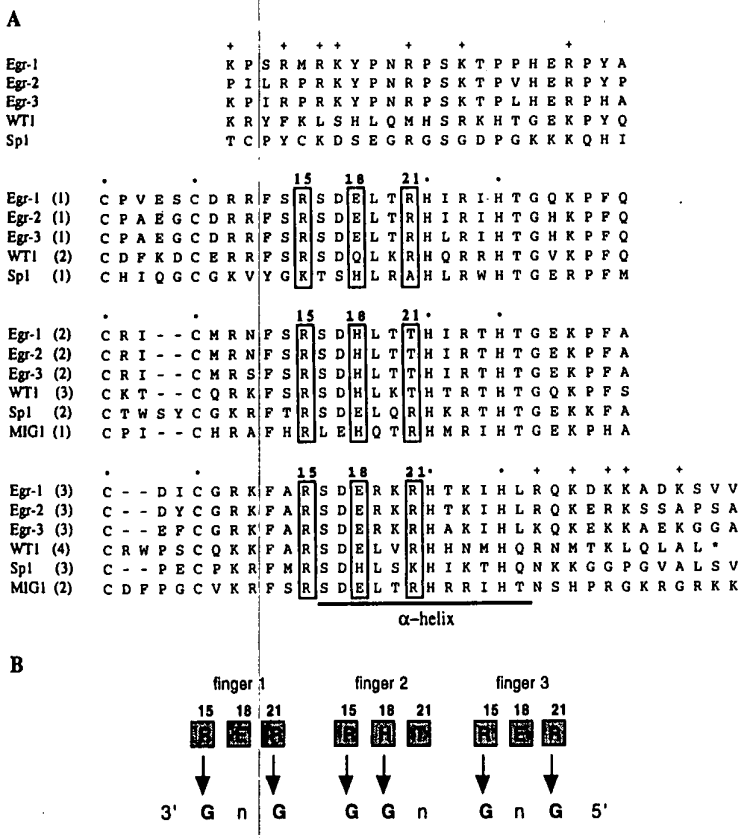


FIG. 5. DNA-binding domains of the EGR family. (A) Comparison of DNA-binding domains related to Egr-1. Zinc fingers and the flanking sequence of murine Egr-1, human Egr-2, human Egr-3, the Wilms tumor gene *WT1*, Sp1, and the yeast protein MIG1 are aligned for comparison. The position of each finger motif is noted in parentheses. Conserved cysteine and histidine residues are marked (·). The helical region is underlined and the residues important in determining binding specificity are enclosed. Conserved basic residues flanking the zinc-finger domains are denoted (+). (B) Residues determining sequence specificity of Egr-1 binding. [Adapted with permission from R. E. Klevit, *Science* 253, 1367 (1991) (Ref. 65). Copyright 1991 American Association for the Advancement of Science.] The Egr-1 zinc-finger domain interacts with the guanine-rich strand of DNA in an antiparallel manner. Fingers 1 and 3 contact the same 3-bp subsite. Arrows represent specific interactions between arginine or histidine residues and the guanine bases.

This structure showed that each finger has a similar relation to the DNA and interacts primarily with a 3-bp subsite. The α-helix of each finger fits directly into the major groove so that residues in the amino-terminal part of the helix

are available for hydrogen-bonding interactions with the base-pairs. The β-sheet is on the backside of the helix away from the base-pairs; the second β-strand of the sheet contacts the sugar-phosphate backbone, serving to orient the α-helix in relation to the DNA. An arginine immediately preceding the helix makes important DNA contacts as do the second, third, and sixth residues of the α-helices. Each of the hydrogen bonds are made to guanines of the G-rich strand of the DNA. The orientation of the three fingers is antiparallel with respect to the G-rich strand; that is, the 5′-most subsite of the sequence is recognized by the carboxy-terminal finger. In addition, the α-helix lies in the major groove in an antiparallel manner, so that the carboxy-terminal portion of each helix interacts with the 5′-most base of each subsite (*64*).

A select number of residues at defined positions in each finger interact specifically with the DNA. The residue preceding the α-helix and the third and sixth residues of the helix make the specific contacts. In each finger of Egr-1, Arg^{15} precedes the helix and Asp^{17} is the second residue of the helix (Fig. 5A). The arginine hydrogen bonds through its long side-chain with a guanine at the third position of each subsite and is stabilized by the conserved aspartate residue. Thus, the third position, G, is common to each subsite and is recognized in an identical manner (Fig. 5B). The third residue of the helix varies between fingers; Glu^{18} is present in fingers 1 and 3 whereas a histidine is present at the same position of finger 2. The structure solved by Pavletich and Pabo (*64*) shows that these glutamate residues do not contact the DNA. In contrast, the histidine of finger 2 participates in a hydrogen bond with the guanine in the center of its subsite. The sixth residue of the helix, Arg^{21} in fingers 1 and 3, forms a specific bond with the guanine occupying the first position of the subsite. A threonine, which is the sixth residue in the helix of finger 2, is incapable of this interaction (*64*).

In summary, a relatively simple pattern has emerged for Egr-1:DNA recognition from this work (*64*). The common arginine immediately preceding the helix specifies a guanine at the third position of each subsite. The third residue of the helix may contact the middle base of the subsite, and the sixth residue of the helix may contact the first base of the subsite. Moreover, the Egr-1 zinc fingers utilize only arginine or histidine residues to contact guanines; in the absence of these amino acids, such as Glu^{18} in fingers 1 and 3 and Thr^{21} in finger 2, there is no specific recognition of the DNA sequence (*64*).

A complementary analysis (*49*) of variant *Egr-1* binding sites has confirmed the lack of specific interactions with the fourth nucleotide: in gel shift assays, GCG-TGG-GCG competed as well as GCG-GGG-GCG. However, the sequences GAG-GGG-GCG and GCG-GGG-GAG were not efficient competitors, showing that not all nucleotides are permissible at positions 2

and 8 (*49*). Although no specific contacts were observed at these positions in the Egr-1:DNA cocrystal (*64*), it is possible that substitution of the bulkier adenosine for cytosine is disruptive at these positions.

These results (*64*, *66*) emphasize the modularity of the zinc-finger motif in which each zinc-binding domain recognizes a three-nucleotide sequence. In particular, an implicit assumption has been that each finger makes an equal contribution to the overall affinity of binding. A complementary *in vivo* mutational analysis of the Egr-1 zinc-finger domain hints that each finger may not make the same contribution to binding. Specifically, many more DNA-binding impaired mutants with alterations in the second finger rather than in the first or third can be recovered (*67*). Moreover, the two His-Cys links connecting the finger motifs also showed a disparity in the number of DNA-binding mutants recovered. The second linker was mutated 17 times whereas the first was altered three times, suggesting that the linkers may not play identical roles in orienting the fingers (*67*).

The recognition code outlined by the crystallographic studies (*66*) indicates similar interactions for all three fingers of Egr-1 and implies that other Cys_2 His_2 zinc-finger proteins will use residues at analogous positions to make their base-specific contacts. Studies with the *Drosophila* finger protein Tramtrack reveal an extension to the formula derived from Egr-1 DNA–protein interactions whereby residues at three positions determine DNA binding specificity. The first finger of Tramtrack uses an additional amino-acid contact to recognize its DNA binding site (*68*). In conclusion, the model developed from Egr-1 studies will generalize to some other zinc-finger proteins, but it does not describe the complete repertoire of all possible protein–DNA contacts in Cys_2 His_2 zinc-finger proteins.

C. Structure–Function Analysis

1. Defining the Egr-1 *Trans*-activation Domains

Definition of a DNA-binding site for Egr-1 set the stage for assessing whether Egr-1 could regulate transcription through the GCG-GGG-GCG sequence. Data from transient transfection assays shows that Egr-1 can activate a minimal promoter with multiple *Egr-1* binding sites 10-fold in a dose-dependent manner (*62*, *69*). Like classical transcription factors, the organization of Egr-1 is modular in nature, with functional domains that are structurally independent and able to confer activity on heterologous proteins. We and others have used deletion analysis and gene fusions to dissect the functional domains of Egr-1, delineating modular activation, repression, and nuclear localization activities.

Deletion analysis of murine Egr-1 indicates that the extensive serine- and threonine-rich N-terminal domain is a robust transcriptional activator. A

constructive approach shows that several Egr-1 activation sequences are independent domains capable of functioning in a heterologous context when fused to the DNA-binding domain of the yeast factor GAL4. Residues 3–281, or subregions from 3 to 138 or 138 to 281, activate transcription 100-fold as GAL4 fusions (*70*). Deletion analysis of the rat homolog of Egr-1 further suggests that residues 13–38 and 223–264 may be most important for the activation function (*57*). The N-terminal domain is 30% serine/threonine/-tyrosine rich over a span of 180 residues; the large size of the activation domain may contribute to its potency relative to the smaller, previously described serine/threonine-rich activator Pit-1/GHF-1 (*71*). Moreover, the *trans*-activation domain is impervious to mutation in that substantial deletions in the extensive N-terminal domain do not destroy transcriptional activity. Finally, work from several laboratories maps a weak trans-activation function to the C-terminus of Egr-1, which contains the octapeptide repeats reminiscent of the phosphorylated YSPTSPS reiterations in the carboxy-terminal domain of RNA polymerase II (*57*, *59*, *70*).

2. Localization of an Egr-1 Repression Domain

An unexpected result of deletion analysis is that a small internal deletion immediately 5′ of the zinc-finger domain (Δ284–330) enhances trans-activation some fivefold in HeLa cells. Western-blotting and gel-shift analyses showed that this superactivation cannot be explained simply by over-expression or enhanced DNA binding of the deletion derivative relative to full-length Egr-1. The superactivation observed with Δ284–330 is consistent with the loss of a region important for repression or for negatively regulating the trans-activation function of Egr-1. Further experiments have shown that *Egr-1* encodes a portable repression domain. Initial work demonstrated that a domain of 34 amino acids (281–314) can repress transcription 7- to 10-fold when fused to the GAL4 DNA-binding domain and assayed for effect on a reporter with five GAL4 binding sites. Repression by this compact domain was dependent on a DNA binding anchor (*70*). A further definition of the essential region showed that residues 281–304 repress and that residues 290–314 are inactive (*72*). This domain, highly conserved throughout vertebrate evolution (*55*), represents a novel motif distinct from the previously described alanine- and glycine-rich repression module in Krüppel (*73*, *74*); the hydrophobic and proline-rich Even-skipped repressor (*75*); the glutamine-, alanine-rich factor Dr1; and the proline-, glycine-rich repressor of WT1 (*76*). In the Egr-1 repression domain, depicted in Fig. 5A, 7 of 24 residues are serine or threonine. In light of the fact that Egr-1 is known to be phosphorylated (*14*, *50*, *60*, *61*), this raises the question of whether the Egr-1 repression function may be regulated by this modification (see below).

Repression by Egr-1 may involve an interaction with a cellular factor. A

competition assay showed that overexpression of Egr-1 amino acids from 266 to 301 results in a dramatic increase in activation from an Egr-1 molecule whose DNA-binding domain has been replaced with that of GAL4 (57). These results suggest that the region 266–301 is sufficient for an interaction with a titratable cellular factor that normally inhibits Egr-1 activity. A single isoleucine-to-phenylalanine substitution at position 290 renders the 266–301 domain nonfunctional. As predicted, this $Ile^{290}Phe$ mutation in the context of the native Egr-1 protein results in dramatic superactivation such that this variant activates about 15 times better than wild-type Egr-1. It is suggested that the cellular factor that interacts through this domain is present in a wide variety of mammalian cells, although apparently not in *Drosophila* Schneider cells because there is no superactivation in this cell type (57).

Elucidation of the mechanism of *Egr-1* repression has begun with the definition of the minimal promoter elements required. Initial work had demonstrated repression with an *Egr-1/GAL4* chimera on a reporter containing a portion of the thymidine kinase promoter with multiple protein-binding elements in addition to a TATA box. However, both *in vivo* and with an *in vitro* transcription assay using bacterially expressed fusion proteins, minimal promoter constructs containing only a TATA or initiator element in addition to binding sites to direct the *Egr-1/GAL4* chimera are sufficient for repression (72). Although these observations suggest that *Egr-1* repression is mediated by some type of interaction with the basal transcription machinery, preliminary experiments indicate that *Egr-1* does not directly bind to either TBP, TFIIB, or TFIIE *in vitro* (77). Therefore, the *Egr-1* repression domain may bind to one of the many other proteins involved in complex formation or to an associated protein, presumably the widely expressed cellular factor titrated by Russo *et al.* (57).

The compact Egr-1 repressor is serine- and threonine-rich, and in particular Thr-289 has homology to known PKC phosphorylation sites (Fig. 5A). Phosphorylation is clearly not required for repression, because bacterially expressed Egr-1 efficiently represses transcription *in vitro* (78). This work is consistent with the suggestion that phosphorylation inactivates the *Egr-1* repression domain, preventing an interaction needed for the transcriptional inhibition. Importantly, an Ile-to-Phe mutation at the position analogous to Egr-1 residue 290 in the PKC substrate neurogranin makes it a better substrate for the kinase (79). The corresponding mutation in Egr-1, which may similarly promote phosphorylation on Thr-289, renders the repression domain nonfunctional (57). The role of phosphorylation may therefore be to enhance the ability of Egr-1 to work as an activator, by muting its repression function.

Egr-1 is one of only a small number of factors that contain modular domains capable of regulating transcription both positively and negatively.

Other examples include the *Drosophila* factor Krüppel (74), YY1/NF-E1/δ (reviewed in 80), and the immediate-early factors Fos and Jun (81). This work provocatively suggests that native Egr-1 may be a bifunctional protein, capable of alternatively activating or repressing transcription. Such a property may be common to immediate-early genes to allow for versatility of effector functions. Posttranslational modifications as discussed above can be envisioned to enable complex factors such as these to regulate transcription either positively or negatively. In the case of Egr-1, we can speculate that Egr-1 may either activate or repress transcription, depending on whether it is induced in response to positive growth or to differentiation cues, or that Egr-1 may activate and repress multiple target genes depending on their promoter context, thereby mediating multiple transcriptional effects in response to a single inducing agent. HL-60 cell differentiation by phorbol may exemplify the latter type of bimodal Egr-1 function. Because Egr-1 expression both promotes macrophage differentiation and prevents granulocytic differentiation, the bifunctional role of Egr-1 may be to stimulate genes essential for macrophage differentiation while repressing genes required for specialized granulocytic functions.

3. Mapping the Egr-1 Nuclear Localization Signal

Consistent with its role as a transcriptional regulator, Egr-1 has been shown by several groups to be localized in the nucleus (50, 60, 61). Small molecules and proteins less than 40–60 kDa may passively diffuse across the nuclear pores into the nucleus, whereas larger proteins are targeted to the nucleus by an active, two-step process. The first step is a rapid, signal-dependent binding to the nuclear pore periphery, and the second step is a slower, ATP- and temperature-dependent translocation across the pore. In a number of nuclear proteins, the signal that specifies nuclear localization (NLS) is generally a short stretch of 8–10 amino acids characterized by basic residues as well as proline (reviewed in 82 and 83).

In Egr-1, basic residues cluster only in the three zinc fingers and adjacent sequences (Fig. 5), hinting that the karyophilic signal of Egr-1 resides here. Using subcellular fractionation/Western analysis or immunocytochemistry to analyze deletion derivatives of Egr-1, we have demonstrated that ΔN314 and ΔC430 are properly targeted to the nucleus, whereas ΔC314 is cytoplasmic. From these results, amino acids 315 to 429, encoding the three zinc fingers and adjacent basic sequences, appear essential for proper nuclear targeting. These results agree with early suggestions that the C-terminus of Egr-1 is required for nuclear localization (60).

A series of fusions of segments of Egr-1 to the large bacterial protein β-galactosidase were further used to show that the zinc-finger domain itself cannot function as an NLS. However, the zinc fingers in conjunction with

the 5′ basic sequence 315–330, but not the 3′ basic sequence, were sufficient to target the bacterial protein β-galactosidase to the nucleus. This 5′ basic stretch of residues 315–330, KPSRMRKYPNRPSKTP, is shared by other members of the EGR family, Egr-2 and Egr-3, which have conserved DNA-binding domains but generally diverge outside this region (Fig. 4A). Additional analyses showed that the entire zinc-finger domain is not required: either finger 2 or 3, yet not finger 1, could work with the 5′ basic sequence to form a bipartite NLS (*70*). Precedents for the incorporation of nuclear targeting signals within a DNA-binding domain include Fos (*84*); the progesterone receptor, in which the second finger but not the first functions as an NLS (*85*); GAL4 (*83*); and the homeodomain proteins α2 and Pit-1/GHF-1 (*71*, *86*). Egr-1 may be a prototypical Cys_2His_2 zinc-finger protein whose DNA-binding and nuclear localization functions have coevolved as a composite domain rich in basic residues.

Other bipartite nuclear localization signals with two basic regions separated by a short variable spacer have been characterized in nucleoplasmin (*87*); SW15 (*88*); the *Xenopus* protein N1 (*89*); the steroid hormone receptors; and polymerase basic protein 1 of influenza virus (*90*). In addition, discontinuous nuclear targeting signals are found in adenovirus DNA-binding protein (*91*) and the yeast repressor α2, which has two nonhomologous signals, i.e., a basic NLS found at the N terminus as well as a signal located in the homeodomain (*86*, *92*). In these proteins, as in Egr-1, the essential domains are discontinuous in the primary sequence, and it has been suggested that the two parts of the signal may mediate separate steps in nuclear accumulation (*86*). Several Egr-1–β-galactosidase mutants containing the 5′ basic sequence (but neither finger 2 nor 3 intact) and showing staining ringing the nucleus may contain the portion of the signal for binding to, but not translocation across, the nuclear pore (*70*).

Each of the assays used to define the Egr-1 NLS measured the equilibrium nuclear/cytoplasmic distribution of protein. Future kinetic analyses may reveal additional sequences required for prompt nuclear localization. Although a signal of seven predominantly basic amino acids suffices for the nuclear accumulation of SV40 T antigen over a period of hours, a more extensive sequence resulted in nuclear targeting within minutes (*93*). Serum-dependent nuclear import has been described for the immediate-early transcription factors c-Fos and, reportedly, c-Jun (*94*). Although Egr-1 is clearly nuclear in serum-stimulated or exponentially growing cells (maintained in 10% calf serum), staining of Egr-1 derivatives or fusion proteins in serum-starved cells should be examined to assess the possibility of conditional nuclear localization.

In conclusion, deletion analysis and Egr-1–β-galactosidase fusions demonstrate that nuclear localization of Egr-1 requires a bipartite signal consist-

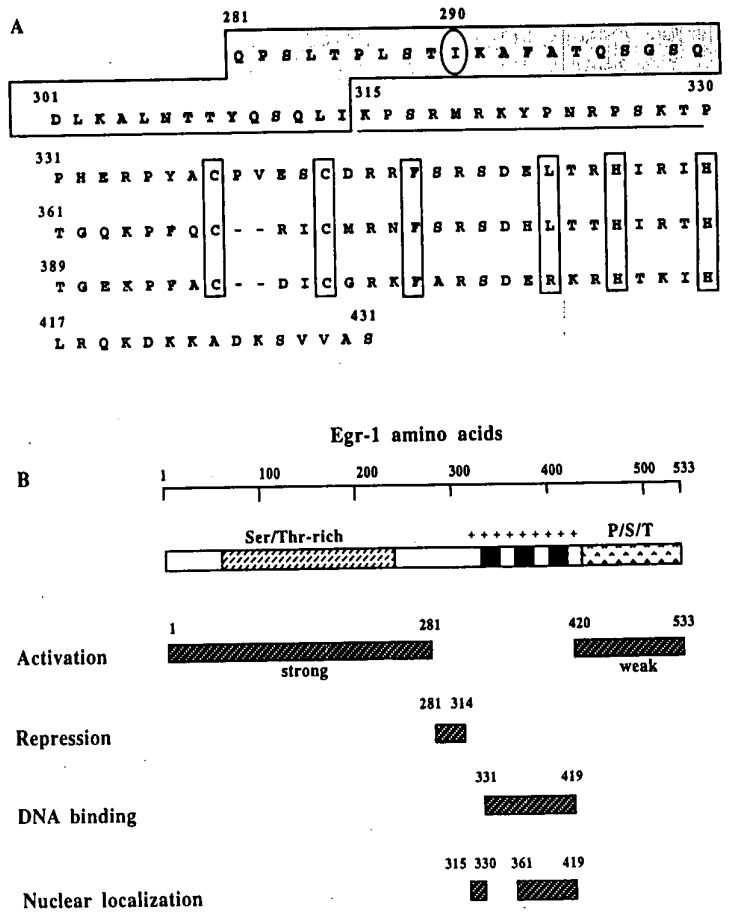


FIG. 6. Summary of Egr-1 domains (modified from *70*). (A) Sequence of Egr-1 repression domain and zinc fingers. The repression domain is shaded and the 5′ basic region involved in nuclear localization is underlined. The threonine residue whose phosphorylation may prevent repression is circled. The three zinc fingers of Egr-1 are aligned for comparison, with residues conserved among Cys_2His_2 zinc fingers enclosed. (B) Functional domains of Egr-1. The serine-/threonine-rich N-terminus of Egr-1 is shown. The basic region of Egr-1 is indicated (+++). Each zinc finger is designated by a black bar and the proline/serine/threonine-rich C-terminal domain (P/S/T) is indicated. Residues 3–281 activate transcription 100-fold and the C-terminus of Egr-1 (residues 420–533) encodes a weaker *trans*-activation function. Amino acids 281–314 suffice to act as a repressor of transcription when fused to a heterologous DNA-binding domain. The DNA-binding activity of Egr-1 has been mapped to amino acids 331–419. The NLS of Egr-1 is bipartite: a basic region (amino acids 315–330) and part of the zinc-finger domain suffice to target Egr-1 to the nucleus.

ing of basic residues 315–330, which flank the zinc-finger domain, in addition to sequences within fingers 2 or 3. These results are notable in light of the fact that relatively few Cys_2His_2 zinc-finger proteins have been characterized with respect to their requirements for nuclear targeting. The incorporation of an NLS within or adjacent to the DNA-binding domain is suggestive of a conserved composite motif in Cys_2His_2 zinc-finger transcription factors (see Fig. 6). Finally, Egr-1 is a member of a small class of proteins that have bipartite nuclear localization signals in which the essential subdomains are separated by more than a few amino acids.

D. Targets of Egr-1 Regulation

1. Genes Regulated in the Context of Cellular Proliferation

Consistent with its induction by mitogenic cues and during terminal differentiation in a few cell types, Egr-1 may bind and regulate genes involved in mitosis or needed for specialized cell functions. The universality of *Egr-1* expression in response to growth signals suggests that genes downstream of *Egr-1* in the cascade governing cellular proliferation will be widely expressed. Several genes belong to this first class of Egr-1 targets, whose regulation presumably directs a cellular response to growth induction.

The expression of the thymidine kinase (*tk*) gene peaks during late G_1, after *Egr-1* induction, kinetics consistent with regulation by *Egr-1*. Enzymes such as thymidine kinase, integral to the biosynthesis of DNA, are regulated depending on the growth state of the cell, and as such thymidine kinase represents a physiologically relevant target for Egr-1. The use of specific α-Egr-1 antiserum (*95*) has demonstrated that Egr-1 is a component of the *tk* promoter-binding complex derived from serum-stimulated nuclear extract. Second, transient transfections in CV-1 cells show that Egr-1 activates a reporter driven by a *tk* promoter fragment from −174 to +159. Egr-1 activation appears to work through a lower affinity binding site, CCG-TGG-GTG. However, it should be noted that because *tk* is also expressed highly in actively cycling cells (in the absence of Egr-1 induction), high-level expression of *tk* apparently does not require Egr-1.

A second target for Egr-1 may be the PDGF A chain, a potent mitogen for cells of mesenchymal origin. PDGF-A is also found at high levels in a number of transformed cell lines. In normal cultured cells, levels of PDGF-A mRNA rise in response to growth factors or cytokines, but peak later than Egr-1 induction. A region of hypersensitivity to the single-strand-specific nuclease S1 in the 5′ untranslated region of PDGF-A that may be involved in regulating transcription of this growth factor has recently been defined (*96*).

Gel-shift competitions with purified Egr-1 showed that this homopurine/homopyrimidine site competes as well as the Egr-1 consensus. Although the S1-sensitive sequence GAG-GAG-GAG-GAGGA deviates at only the underlined position from the contacts determined by crystallography to be important for binding (*64*), the high affinity of the S1-sensitive site is surprising considering that previous studies have shown that GAG in the first or third subsite is not optimal for Egr-1 binding (*97*). Nevertheless, this homopurine/homopyrimidine sequence may be of widespread importance, because similar motifs derived from the promoters of other growth-related genes, such as the epidermal growth factor receptor, the insulin receptor, c-Ki-*ras*, c-*myc*, and TGF-β3, are also good competitors of Egr-1 binding (*96*). Future studies will determine if these provocative *in vitro* studies are of physiological significance by assessing whether Egr-1 can regulate transcription of the PDGF-A gene through this variant motif.

A third Egr-1 target in primary fetal astrocytes may be bFGF. An antisense oligomer to *Egr-1* blocks bFGF induction following addition of a mitogen, ET-3 (*23*).

2. Genes Regulated in the Context of Cell Differentiation

Expression of the myosin heavy chain α gene (α-*MHC*) and *Egr-1* are coregulated in serum-deprived primary cultures of cardiac myocytes stimulated with serum and when the embryonal carcinoma cell line P19 differentiates into cardiac cells in response to dimethyl sulfoxide, prompting investigation of α-MHC as a target of Egr-1 regulation. A CAT reporter containing 1.7 kb of the α-myosin heavy chain promoter is activated 10-fold by an *Egr-1* expression vector in transfected primary cultures of fetal rat cardiac myocytes. Northern analysis shows the endogenous α-*MHC* gene is also stimulated three- to fourfold by Egr-1 (*98*). Induction of α-*MHC* in response to Egr-1 was observed in the myogenic Sol8 cell line, but not in NIH3T3 fibroblasts, suggesting tissue-specific induction; α-*MHC* expression was unchanged in response to Egr-1 in another muscle cell line, L_6E_9, showing that Egr-1 is not sufficient for the *MHC* gene activation. The region of the rat α-*MHC* promoter that is Egr-1 responsive to a segment from −1698 to −1283 has been delimited (*98*). A potential Egr-1 binding site GTG-GGG-GTG is located within this promoter fragment, but has not yet been shown to be the functional element (*98*). In light of the study showing that Egr-1 translation is blocked during Sol8 differentiation in response to insulin (*20*), Egr-1 protein levels in cardiac myocytes remain to be analyzed.

A functional role for Egr-1 in adrenergic differentiation, suggested by high-level expression in the rat adrenal gland and in PC12 cells, may be the

regulation of phenylethanolamine *N*-methyltransferase (PNMT), the adrenal enzyme that converts norepinephrine to epinephrine. *In vivo*, neural stimulation causes an increase in Egr-1 protein in the adrenal medulla corresponding to a rise in PNMT expression in the same cell type (*99*). Transient transfections in the highly transfectable PC12 subline RS1 reveal that Egr-1 can modestly stimulate (fourfold) a PNMT reporter with 442 bp of 5′ sequence. This region includes two potential Egr-1 binding sites, an optimal consensus sequence at −165 and a proximal site GCG-GGG-GGG at −45. Cold competition experiments show that this 8 of 9 match to the optimal Egr-1 consensus is a weak but specific competitor (*99*).

It has been postulated that Egr-1 negatively regulates the widely expressed adenosine deaminase gene. ADA has a (G + C)-rich promoter, lacking TATA and CCATT box elements, typical of classical housekeeping gene promoters. As discussed above, Egr-1 and Sp1 bind to distinct, although similarly (G + C)-rich, DNA-binding sites (*97*). Notably, deletion analysis of the ADA promoter reveals a *cis*-acting repressor element that maps to an Egr-1 site. Mutations that destroy Egr-1 binding but do not affect the Sp1 site in the 13 bp overlapping Egr-1/Sp1 motif GCG-TGG-GCG-GGGC result in a 15-fold enhancement in promoter activity. *In vitro*, Egr-1 and Sp1 protect overlapping segments of this complex 13-bp sequence. One hypothesis is that Egr-1 negatively regulates ADA transcription by competitively occupying the motif and displacing Sp1. Alternatively, Egr-1 may repress the ADA promoter by an active mechanism independent of the Sp1, also consistent with results described above. Evidence for this proposal is that even in the absence of an Sp1 site, mutation of the Egr-1 motif results in higher promoter activity. Future studies with varying ratios of Egr-1 and Sp1 expression vectors as well as experiments addressing the issue of whether the Egr-1 DNA-binding domain is sufficient for the negative regulation will be informative.

The definition of a consensus binding site for Egr-1 has propelled investigations to identify the genes that Egr-1 binds and regulates. The *tk* gene represents a physiologically relevant target for Egr-1 in the context of cell growth. The induction of the *tk* gene subsequent to the Egr-1 serum response, the ability of Egr-1 to bind to a site in the *tk* 5′ sequence, and transcriptional activation of the *tk* promoter by Egr-1 in transient transfections all support the idea that thymidine kinase is an important Egr-1 target. A second Egr-1 target may be the mitogen PDGF, because Egr-1 can bind to a site in the PDGF-A gene (*100*). Other potential Egr-1 target genes are clearly not relevant to cellular proliferation. The expression pattern of Egr-1 in the adult animal as well as its induction during terminal differentiation in some cell types suggest that Egr-1 plays a role in specialized cells that is distinct from its function during the G_0 to G_1 transition. In cardiac cells, the

endogenous α-myosin heavy chain gene or a transfected construct containing the α-*MHC* promoter is stimulated by Egr-1 (*98*). And in adrenal cells, Egr-1 can regulate the phenylethanolamine *N*-methyltransferase gene, supporting a role for Egr-1 in adrenergic differentiation (*99*). Additional targets of Egr-1 regulation in other differentiated cell types, for example, specific to osteoblasts or to macrophages, remain to be identified.

VI. *In Vivo* Role of Egr-1

The challenge remaining in current Egr-1 research is to relate correlative expression data and *in vitro* studies to a biological role for Egr-1. In a few instances, overexpression or antisense analyses have shown a phenotype for Egr-1. These studies have focused on differentiated cell types; despite the abundance of data showing Egr-1 induction by mitogenic signals, a role for Egr-1 in cell growth/division remains to be established. These phenotypic analyses are complicated by potential functional redundancy contributed by related members of the EGR family (see Section VII). With virtually identical DNA-binding domains, the expression of related family members may serve to mask a phenotype in Egr-1 loss-of-function experiments.

A clear-cut biological role for Egr-1 has been demonstrated in three systems. As discussed above, antisense oligomers preventing Egr-1 expression in myeloid cells block macrophage differentiation. Further, constitutive Egr-1 expression restricts the potential of HL-60 cells, rendering them incapable of differentiation along the granulocyte lineage (*32*). A second phenotype for Egr-1 involves its role as a positive regulator of astrocyte proliferation as discussed earlier (*23*). A third system in which Egr-1 plays a causal role involves the hypertrophic growth of cardiac myocytes in response to endothelin-1. *Egr-1*, as a gene rapidly induced by endothelin, was proposed to mediate cardiac hypertrophy. It has been definitively shown that endothelin-1-induced hypertrophic growth in adult rat cardiomyocytes, as assayed by increased protein synthesis, is blocked by oligomers complementary to the *Egr-1* message (*101*). Additional phenotypes for Egr-1 await experimentation in other systems. Perhaps the most exciting unanswered question is whether *Egr-1* functions as a cellular proto-oncogene in a manner analogous to c-*fos*.

VII. *Egr-1* Is Part of a Gene Family, Including the Wilms Tumor Suppressor Gene *WT1*

Egr-1 shares a highly conserved domain, encoding the three zinc-finger motifs, with several other immediate-early genes as well as genes that func-

tion in unrelated contexts. Zinc-finger proteins of the type first described for TFIIIA contains invariant residues, including conserved cysteines and histidines, but also include nonconserved residues that presumably dictate the specificity of binding. *Egr-2/Krox20 (102, 103)*, *Egr-3 (69)*, and *Egr-4/NGFI-C/pAT133 (69, 104, 105)* encode proteins with zinc-finger domains virtually identical to that of Egr-1. The Egr-1 zinc-finger domain is over 95% identical to that of Egr-2 and 91% identical to that of Egr-3 at the amino-acid level. Most of the changes are conservative substitutions, and residues important in determining the sequence specificity of binding are absolutely conserved (Fig. 5A). The homology extends to adjacent basic sequences but drops abruptly outside this region. *Egr-2/Krox20* and *Egr-3* are strikingly induced by growth factors whereas *Egr-4/NGFI-C/pAT133* is more weakly inducible (*105*). The expression of *Egr-2/Krox 20*, restricted to the nervous system during mouse embryogenesis, generates a segment-specific pattern in the developing hindbrain (*69, 103, 106*). Importantly, disruption of *Egr-2/Krox20* by homologous recombination in the mouse results in postnatal death of the animal, with anatomical analysis showing severely reduced or absent rhombomeres 3 and 5 in the hindbrain; (*107*). Finally, *Egr-2/Krox20* brain expression is also transiently activated by electroconvulsive shock treatment, D1 dopamine receptor activation, and opiate withdrawal, in a pattern similar to that noted for *Egr-1/Zif 268 (41)*.

The Wilms tumor suppressor gene *WT1*, implicated in the genesis of this pediatric kidney malignancy, has four zinc fingers, three of which are highly homologous (67% identical) to the Egr-1 zinc-finger domain (*108, 109*). The WT1 protein binds to the EGR consensus binding sequence GCG-GGG-GCG but with lower affinity. Moreover, the first finger of WT1 and the presence of KTS (an alternatively spliced variant) between fingers 3 and 4 dictate other sequence requirements for DNA binding (55). The mammalian activator Sp1 also has three related zinc fingers, with finger 2 most similar to EGR fingers 1 and 3 (*110*). The EGR family of proteins is also distantly related to MIG1, a yeast protein that responds to glucose repression (*111*). As suggested by the homology of the zinc-finger motifs, the sequences recognized by the EGR proteins, WT1, and Sp1 are related. Interestingly, flanking (A + T)-rich sequences play critical roles in target site recognition by MIG1. These flanking sequence preferences may reflect local DNA binding (*112*).

VIII. Conclusion and Future Perspectives

The genomic response of a cell to changes in its extracellular environment includes the induction of immediate-early transcription factor genes.

The most extensively characterized members of this group are the *fos*, *jun*, and *Egr* family members. Their discovery has allowed delineation of the "proximal" events from cell surface to nucleus that induce them: definition of intracellular signaling pathways and downstream promoter elements they target. More recent efforts have focused on events "distal" to transcription factor gene induction: characterization of the proteins involved, their interactions with each other, definition of the target DNA sequences to which they bind, structure–function analyses, negative regulation following induction, and other forms of cross-talk between these family members. Collectively, therefore, these investigations have enhanced our knowledge of signal transduction pathways and general mechanisms of transcriptional activation and repression, and protein-DNA interactions. The most important critical questions for future analysis involve the further identification of phenotypes, either by ectopic overexpression or by "underexpression" using dominant negative, antisense, or homologous recombination methodologies. Unfortunately, however, many phenotypes may be masked by redundant pathways. Nevertheless, a search for such systems will be critical to provide the substrate by which to characterize suitable physiological target genes for immediate-early transcription factor action.

ACKNOWLEDGMENT

This work was partially supported by NIH Grant CA40046 to V.P.S.

REFERENCES

1. D. Denhardt, D. Edwards and C. Parfett, *BBA* 865, 83 (1986).
2. H. Herschman, *TIBS* 14, 455 (1989).
3. H. Herschman, *ARB* 60, 281 (1991).
4. L. Lau, *Curr. Opin. Cell Biol.* 2, 280 (1990).
5. V. P. Sukhatme, S. Kartha, F. Toback, R. Taub, R. Hoover and C.-H. Tsai-Morris, *Oncogene Res.* 1, 343 (1987).
6. V. Sukhatme, X. Cao, L. Chang, C.-H. Tsai-Morris, D. Stamenkovich, P. Ferreira, D. Cohen, S. Edwards, T. Shows, T. Curran, M. Le Beau and E. Adamson, *Cell* 53, 37 (1988).
7. S. Suggs, J. Katzowitz, C. Tsai-Morris and V. Sukhatme, *NARes* 18, 4283 (1990).
8. J. Milbrandt, *Science* 238, 797 (1987).
9. L. Lau and D. Nathans, *PNAS* 84, 1182 (1987).
10. R. Lim, B. Varnum and H. Herschman, *Oncogene* 1, 263 (1987).
11. D. Simmons, D. Levy, Y. Yannoni and R. Erikson, *PNAS* 86, 1178 (1989).
12. J. Wright, K. Gunter, H. Mitsuya, S. Irving, K. Kelly and U. Siebenlist, *Science* 248, 588 (1990).
13. P. Lemaire, O. Revelant, R. Bravo and P. Charnay, *PNAS* 85, 4691 (1988).

14. B. Christy, L. Lau and D. Nathans, *PNAS* **85**, 7857 (1988).
15. S. Qureshi, C. Joseph, M.-H. Rim, A. Maroney and D. Foster, *Oncogene* **6**, 995 (1991).
16. K. Alexandropoulos, S. A. Qureshi, M. Rim, V. P. Sukhatme and D. A. Foster, *NARes* **20**, 2355 (1992).
17. V. Seyfert, V. Sukhatme and J. Monroe. *MCBiol* **9**, 2083 (1989).
18. V. Seyfert, S. McMahon and W. Glenn. *Science* **250**, 797 (1990).
19. H. D. Rupprecht, P. Dann, V. P. Sukhatme, R. B. Sterzel and D. L. Coleman, *Am. J. Physiol.* **263**, F623 (1992).
20. B. Wollnik, C. Kubisch, A. Maass, H. Vetter and L. Neyses, *BBRC* **194**, 642 (1993).
21. X. M. Cao, G. R. Guy, V. P. Sukhatme and Y. H. Tan, *JBC* **267**, 1345 (1992).
22. F. Rastinejad and H. Blau, *Cell* **72**, 903 (1993).
23. R.-M. Hu and E. R. Levin, *J. Clin. Invest.* **93**, 1820 (1994).
24. J. Milbrandt, *Neuron* **1**, 183 (1988).
25. M. Watson and J. Milbrandt, *Development* **110**, 173 (1990).
26. A. McMahon, J. Champion, J. McMahon and V. Sukhatme, *Development* **108**, 281 (1990).
27. S. Vainio, I. Karavanova, A. Jowett and I. Thesleff, *Cell* **75**, 45 (1993).
28. T. Darland, S. Samuels, V. P. Sukhatme and E. Adamson. *Oncogene* **6**, 1367 (1991).
29. L. Suva, M. Ernst and G. Rodan. *MCBiol* **11**, 2503 (1991).
30. S. Kharbanda, E. Rubin, R. Datta, R. Hass, V. Sukhatme and D. Kufe. *Cell Growth Differ.* **4**, 17 (1993).
31. S. Kharabanda, T. Nakamura, R. Stone, R. Hass, S. Bernstein, R. Datta, V. P. Sukhatme and D. Kufe, *J. Clin. Invest.* **88**, 571 (1991).
32. H. Nguyen, B. Hoffman-Liebermann and D. Leibermann, *Cell* **72**, 197 (1993).
33. M. Le Beau, R. Espinosa III and W. Neuman, *PNAS* **90**, 5484 (1993).
34. J. V. Bonventre, V. P. Sukhatme, M. Bamberger, A. J. Ouellette and D. Brown. *Cell Reg.* **2**, 251 (1991).
35. D. Hallahan, V. P. Sukhatme, M. Sherman, S. Virudachalam, D. Kufe and R. R. Weichselbaum, *PNAS* **88**, 2156 (1991).
36. J. Morgan and T. Curran, *Trends Neurosci.* **12**, 459 (1989).
37. D. Bartel, M. Sheng, L. Lau and M. Greenberg, *Genes Dev.* **3**, 314 (1989).
38. K. Rosen, M. McCormack and L. Villa-Komaroff. *PNAS* **89**, 5437 (1992).
39. P. Worley, B. Christy, Y. Nakabeppu, R. Bhat, A. Cole and J. Baraban, *PNAS* **88**, 5106 (1991).
40. A. Cole, D. Saffen, J. Baraban and P. Worley, *Nature NB* **340**, 474 (1989).
41. R. Bhat, P. Worley, A. Cole and J. Baraban, *Brain Res. Mol. Brain Res.* **13**, 263 (1992).
42. T. Herdegen, K. Kovary, J. Leah and R. Bravo, *J. Comp. Neurol.* **313**, 178 (1991).
43. G. Jamieson, R. Mayforth, M. Villereal and V. P. Sukhatme, *J. Cell Physiol.* **139**, 262 (1989).
44. V. Seyfert, S. McMahon, W. Glenn, X. Cao, V. Sukhatme and J. Monroe, *J. Immunol.* **145**, 3647 (1990).
45. S. A. Qureshi, M. Rim, J. Bruder, W. Kolch, U. Rapp, V. P. Sukhatme and D. A. Foster, *JBC* **266**, 20594 (1991).
46. C.-H. Tsai-Morris, X. Cao and V. P. Sukhatme, *NARes* **16**, 8835 (1988).
47. P. Changelian, P. Feng, T. King and J. Millbrandt, *PNAS* **86**, 377 (1989).
48. R. Treisman, *Cell* **46**, 567 (1986).
49. B. Christy and D. Nathans, *MCBiol* **9**, 4889 (1989).
50. X. Cao, R. Koski, and A. Gashler *et al*, *MCBiol* **10**, 1931 (1990).
51. D. Gius, X. Cao, F. Rauscher, D. Cohen, T. Curran and V. Sukhatme, *MCBiol* **10**, 4243 (1990).
52. S. Qureshi, X. Cao, V. Sukhatme and D. Foster, *JBC* **266**, 10802 (1991).

53. R. Datta, E. Rubin, V. Sukhatme, S. A. Qureshi, R. Weichselbaum and D. W. Kufe, *PNAS* **89**, 10149 (1992).
54. J. Miller, A. McLachlan and A. Klug, *EMBO J.* **4**, 1609 (1985).
55. I. Drummond, H. Rupprecht, P. Rohwer-Nutter, J. Lopez-Guisa, S. Madden, F. Rauscher III and V. Sukhatme, *MCBiol* **14**, 3800 (1994).
56. J. Lanfear, T. Jowett and P. Holland, *BBRC* **179**, 1220 (1991).
57. M. Russo, C. Matheny and J. Milbrandt, *MCBiol* **13**, 6858 (1993).
58. R. Richards and G. Sutherland, *Cell* **70**, 709 (1992).
59. J. Corden, *TIBS* **15**, 383 (1990).
60. M. Day, T. Fahrner, S. Aykent and J. Milbrandt, *JBC* **25**, 15253 (1990).
61. C. Waters, D. Hancock and G. Evan, *Oncogene* **5**, 669 (1990).
62. P. Lemaire, C. Vesque, J. Schmitt, H. Stunnenberg, R. Fank and P. Charnay, *MCBiol* **10**, 3456 (1990).
63. J. Berg, *Annu. Rev. Biophys. Biophys. Chem.* **19**, 405 (1990).
64. N. P. Pavletich and C. O. Pabo, *Science* **252**, 809 (1991).
65. R. E. Klevit, *Science* **253**, 1367 (1991).
66. J. Nardelli, T. Gibson, C. Vesque and P. Charnay, *Nature NB* **349**, 175 (1991).
67. T. Wilson, M. Day, T. Pexton, K. Padgett, M. Johnston and J. Milbrandt, *JBC* **267**, 3718 (1992).
68. L. Fairall, J. Schwabe, L. Chapman, J. Finch and D. Rhodes, *Nature NB* **366**, 483 (1993).
69. S. Patwardhan, S. Gashler, M. Siegel, L. Chang, L. Joseph, T. Shows, M. Le Beau and V. Sukhatme, *Oncogene* **6**, 917 (1991).
70. A. L. Gashler, S. Swaminathan and V. P. Sukhatme, *MCBiol* **13**, 4556 (1993).
71. L. Theill, J.-L. Castillo, D. Wu and M. Karin, *Nature NB* **342**, 945 (1989).
72. N. Zeleznik-Le, A. Gashler and V. Sukhatme, unpublished (1994).
73. J. Licht, M. Grossel, J. Figge and U. Hansen, *Nature NB* **346**, 76 (1990).
74. P. Zuo, D. Stanojevic, J. Colgan, K. Han, M. Levine and J. Manley, *Genes Dev.* **5**, 254 (1991).
75. K. Han and J. Manley, *Genes Dev.* **7**, 491 (1993).
76. S. Madden, D. Cook and F. I. Rauscher, *Oncogene* **8**, 1713 (1993).
77. N. Zeleznik-Le, V. Sukhatme, R. Drapkins and D. Reinberg, unpublished (1993).
78. N. Zeleznik-Le and V. Sukhatme, unpublished (1993).
79. S.-J. Chen, E. Klann, M. Gower, C. Powell, J. Sessoms and J. Sweatt, *BJ* **32**, 1032 (1993).
80. S. Hahn, *Curr. Biol.* **2**, 152 (1992).
81. C. Abate, D. Luk and T. Curran, *MCBiol* **11**, 3624 (1991).
82. J. Garcia-Bustos, J. Heitman and M. Hall, *BBA* **1071**, 83 (1991).
83. P. Silver, L. Keegan and M. Ptashne, *PNAS* **81**, 5951 (1984).
84. I. Tratner and I. Verma, *Oncogene* **6**, 2049 (1991).
85. A. Guichon-Mantel, P. Lescop, S. Christin-Maitre, H. Loosfelt, M. Perrot-Appanat and E. Milgrom, *EMBO J.* **10**, 3851 (1991).
86. M. Hall, C. Craik and Y. Hiraoka, *PNAS* **87**, 6954 (1990).
87. J. Robbins, S. Dilworth, R. Laskey and C. Dingwall, *Cell* **64**, 615 (1991).
88. T. Moll, G. Tell, U. Surana, H. Robitsch and K. Nasmyth, *Cell* **66**, 743 (1991).
89. J. Kleinschmidt and A. Seiter, *EMBO J.* **7**, 1605 (1988).
90. S. Nath and D. Nayak, *MCBiol* **10**, 4139 (1990).
91. N. Morin, C. Delsert and D. Klessig, *MCBiol* **9**, 4372 (1989).
92. M. Hall, L. Hereford and I. Herskowitz, *Cell* **36**, 1057 (1984).
93. H.-P. Rihs and R. Peters, *EMBO J.* **8**, 1479 (1989).
94. P. Roux, J.-M. Blanchard, A. Fernandez, N. Lamb, P. Jeanteur and M. Piechaczyk, *Cell* **63**, 341 (1990).

95. A. Crozat, G. Molnar and A. B. Pardee, *MCBiol* (1995). Submitted.
96. Z. Wang and T. Deuel, *BBRC* 188, 433 (1992).
97. B. Christy and D. Nathans, *PNAS* 86, 8737 (1989).
98. M. P. Gupta, G. Gupta, R. Zak and V. P. Sukhatme, *JBC* 266, 12813 (1991).
99. S. Ebert, S. Balt, J. Hunter, A. Gashler, V. Sukhatme and D. Wong, *JBC*, 269, 20885 (1994).
100. Z. Wang, S. Madden, T. Deuel and F. Rauscher, *JBC* 267, 21999 (1992).
101. L. Neyses, J. Nouskas and H. Vetter, *BBRC* 181, 22 (1991).
102. P. Chavrier, M. Zerial, P. Lemaire, J. Almendral, R. Bravo and P. Charnay, *EMBO J.* 7, 29 (1988).
103. L. Joseph, M. Le Beau and G. Jamieson, *et al.*, *PNAS* 85, 7164 (1988).
104. S. Crosby, J. Puetz, K. Simburger, T. Fahrner and J. Milbrandt, *MCBiol* 11, 3835 (1991).
105. H.-J. Müller, C. Skerka, A. Bialonski and P. Zipfel, *PNAS* 88, 10079 (1991).
106. D. Wilkinson, S. Bhatt, P. Chavrier, R. Bravo and P. Charnay, *Nature NB* 337, 461 (1989).
107. S. Schneider-Maunoury, P. Topilko and T. Seitandou *et al.*, *Cell*, 75, 1199 (1993).
108. M. Gessler, A. Poustka, W. Cavenee, R. Neve, S. Orkin and G. Bruns, *Nature NB* 343, 774 (1990).
109. K. Call, T. Glaser, C. Ito, A. Buckler, J. Pelletier, D. A. Haber, E. A. Rose, A. Kral, H. Yeger, W. H. Lewis, C. Jones and D. E. Housman, *Cell* 60, 509 (1990).
110. J. Kadonaga, K. Carner, F. Masiarz and R. Tjian, *Cell* 51, 1079 (1987).
111. J. Nehlin and H. Ronne, *EMBO J.* 9, 2891 (1990).
112. M. Lundin, J. Nehlin and H. Ronne, *MCBiol* 14, 1979 (1994).
113. M. H. Little, J. Prosser, A. Condie, P. J. Smith, V. Van Heyningen and N. D. Hastie, *PNAS* 89, 4791 (1992).
114. H. Rupprecht, V. P. Sukhatme, J. Lacy, R. B. Sterzel and D. L. Coleman, *Am. J. Physiol.* 265, F351 (1993).
115. D. L. Coleman, A. H. Bartiss, V. P. Sukhatme, J. Liu and H. D. Rupprecht, *J. Immunol.* 149, 3045 (1992).
116. R. Datta, N. Taneja, V. P. Sukhatme, S. A. Qureshi, R. Weichselbaum and D. W. Kufe, *PNAS* 90, 2419 (1993).
117. L. Neyses, J. Nouskas, J. Luyken, S. Fronhoffs, S. Oberdorf, U. Pfeifer, R. S. Williams, V. P. Sukhatme and H. Vetter. *J. Hypertens.* 11, 927 (1993).
118. D. M. Cohen and S. R. Gullans, *Am. J. Physiol.* 264, F593 (1993).
119. B. A. Haber, K. L. Mohn, R. H. Diamond and R. Taub, *J. Clin. Invest.* 91, 1319 (1993).
120. A. Sachinidis, P. Weisser, Y. Ko, K. Schulte, K. Meyer zu Brickwedde, L. Neyses and H. Vetter, *FEBS Lett.* 313, 109 (1992).
121. J. W. Liu, J. Lacy, V. P. Sukhatme and B. L. Coleman, *JBC* 266, 5629 (1991).

Two New Collagen Subgroups: Membrane-associated Collagens and Types XV and XVIII

TAINA PIHLAJANIEMI AND MARKO REHN

Collagen Research Unit
Biocenter and Department of Medical Biochemistry
University of Oulu
FIN-90220 Oulu, Finland

I. The Collagen Superfamily 226
A. Fibrillar Collagens 228
B. Nonfibrillar Collagens 228
II. Membrane-associated Collagenous Proteins 230
A. Macrophage Scavenger Receptors 232
B. Complement Subcomponent C1q 234
C. Type XVII Collagen 235
D. Type XIII Collagen 237
E. Deliberations on Membrane-associated Collagenous Proteins 246
III. Collagen Types XV and XVIII 248
A. Structural Characteristics of the α1(XV) and α1(XVIII) Chains 248
B. Sequence Homologies between Collagen Types XV and XVIII 250
C. Genes Encoding Collagen Types XV and XVIII 252
D. Tissue Distribution of mRNAs for Collagen Types XV and XVIII 253
E. Variant Type XVIII Collagen Chains Are Homologous with Tissue Polarity Gene Products ("Frizzled" Proteins) 253
F. Deliberations on Collagen Types XV and XVIII 256
IV. Conclusions and Perspectives 257
References 259

The collagens comprise a large family of genetically distinct but structurally related proteins that are found in essentially all connective tissues of most multicellular organisms, being particularly abundant in cartilage, demineralized bone, ligaments, placenta, tendon, skin, and most blood vessels. A prominent function of collagens is to maintain the architecture of tissues and organs and to confer strength on them, but they are also involved in early development and organogenesis, cell attachment, chemotaxis, and filtration through basement membranes.